



## Arquitetura Transformer Aplicada À Detecção De Anomalias Em Pontes

Igor J.S. Ribeiro<sup>1</sup>, Cássio Bragança<sup>2</sup>, Andréia Meixedo<sup>3</sup>, Diogo Ribeiro<sup>4</sup>, Marcelo O. Silva<sup>5</sup>, André W. Brasil<sup>6</sup>, Túlio N. Bittencourt<sup>7</sup>

<sup>1,2,7</sup> Universidade de São Paulo /Departamento de Estruturas e Fundações/igorjsribeiro@usp.br

<sup>3</sup> CONSTRUCT - LESE, Faculdade de Engenharia da Universidade de Porto

<sup>4</sup> CONSTRUCT - LESE, Escola de Engenharia da Politécnica do Porto

<sup>1,5,6</sup> Centro Integrado de Manufatura e Tecnologia (SENAI-CIMATEC)

### Resumo

Tendo em vista a importância das pontes para infraestrutura, novas metodologias são continuamente propostas, visando assegurar desempenho em serviço adequado ao longo de toda vida útil. De particular interesse são as metodologias orientadas por dados, que viabilizam abordagens menos subjetivas, robustas e passíveis de aplicação em cenários reais. Nesse cenário, o presente trabalho visa apresentar um estudo sobre o uso de modelos transformers, no contexto de séries temporais, para detecção de anomalias utilizando uma abordagem baseada em resíduos. A partir do conhecido benchmark da ponte ferroviária KW51, que conta com dados coletados antes e após a alteração global das ligações dos arcos com o tabuleiro, um modelo transformer foi treinado utilizando os dados de aceleração para prever valores, e extrair os resíduos dessa predição. Em sequência, uma análise dos parâmetros de treino é efetuada, delineando convergência e possibilidades de melhoria para generalização da arquitetura. Finalmente, os resíduos são utilizados para rotular os dados provenientes do sensor como alterados ou não. Os resultados demonstram o potencial da arquitetura para aprender contexto a partir de pequenos vetores temporais e ao mesmo tempo delimitam os problemas a serem resolvidos em termos de complexidade computacional para utilização extensiva, apta a detectar influências que se desenvolvem ao longo de maiores sequências.

### Palavras-chave

Transformer – Modelos orientados por dados – Detecção de dano em pontes.

### 1 Introdução

O transporte ferroviário, tanto no Brasil quanto globalmente, tem experimentado uma evolução constante. Para competir com outros modos de transporte, a velocidade e a capacidade de carga dos trens estão em contínuo aumento. No entanto, tal progresso suscita preocupações relativas à segurança, considerando uma infraestrutura viária antiga que não acompanha o desenvolvimento acelerado dos veículos ferroviários. Especialmente no modal ferroviário, pontes e viadutos são componentes altamente críticos, tendo em vista a impossibilidade de se realizar desvios na eventualidade de um colapso e na grande dificuldade que veículos ferroviários têm de transpor aclives. Tudo isso reforça o quanto é crucial monitorar a integridade estrutural de componentes críticos, como pontes e viadutos, para identificar danos de forma precoce e orientar ações de reparo adequadas (Alves et al., 2015; Carnevale et al., 2019).

Tanto no Brasil quanto no mundo, a prática corrente na gestão de ativos ainda se baseia em inspeções periódicas para a detecção de manifestações patológicas executadas conforme as diretrizes da NBR 9452 (2019). Tipicamente, essas inspeções consistem na avaliação visual das estruturas por técnicos especializados, enquanto o emprego de técnicas mais sofisticadas, como instrumentação avançada e ensaios não destrutivos, ainda é pouco difundido. Embora este tipo de gestão de ativos seja empregado há muitos anos, ele é altamente sujeito à subjetividade na avaliação do inspetor, além de enfrentar questões logísticas de acesso aos ativos, entre outros desafios. Ademais, o intervalo entre essas inspeções pode resultar em atrasos significativos na detecção de falhas. Dado este cenário de gestão pouco eficiente atrelado a um aumento cada

vez mais expressivo dos custos associados a uma falha destes ativos os gestores de infraestrutura têm buscado formas mais confiáveis para gerir seus ativos.

Dado este crescente interesse por formas cada vez mais confiáveis de se monitorar os ativos o campo de SHM – *Structural Health Monitoring* tem ganhado cada vez mais força como uma valiosa ferramenta de gestão da integridade de ativos de infraestrutura. Em suma sistemas de SHM baseiam-se em uma avaliação contínua e objetiva de uma estrutura, usualmente feita por meio de sensores, que buscam correlacionar mudanças no comportamento estrutural com a degradação de uma estrutura (Meixedo et al., 2021). Conforme definido por (Farrar & Worden, 2012) a implementação de uma metodologia de SHM reside fundamentalmente em quatro etapas principais: i) Avaliação operacional, ii) Aquisição de dados, iii) Extração de atributos e iv) Discriminação de atributos. Vale destacar que inerente às etapas ii a iv estão procedimentos de modelagem, limpeza, normalização, fusão e compressão de dados.

Dentre todas estas etapas a extração de atributos é a mais crucial do processo, pois os dados brutos não são suficientemente informativos sobre a ocorrência de danos. Historicamente propriedades modais, mais especificamente frequências naturais, foram amplamente utilizadas como indicadores da presença de dano. Entretanto, conforme demonstrado por Meixedo et al (2021) as alterações das repostas estruturais decorrentes de danos, principalmente aqueles em estado inicial são pequenas podendo ser facilmente mascaradas por interferências ambientais como mudanças de temperatura, ventos, chuva, dentre muitas outras. Para além disso, fatores operacionais como variações na velocidade dos trens, carga por vagão e até mesmo o estado de manutenção das suspensões podem impactar nos resultados. Frente a isso, parâmetros modais tem se mostrado insuficientes como indicadores de dano tendo em vista sua forte sensibilidade a estas interferências principalmente variações de temperatura. Dada esta constatação, pesquisadores tem buscado por estratégias alternativas para a extração de atributos mais sensíveis ao dano. Técnicas como transformadas Wavelet (Kankanamge et al., 2020), transformada Hilbert (Chen et al., 2014), funções de resposta em frequência (Ruiz et al., 2024), momentos estatísticos (Finotti et al., 2019), entre outras vem se mostrando promissoras na extração de atributos. Uma outra abordagem que vem galgando aos poucos envolve a utilização de modelos de série temporal. Nesse contexto, modelos autorregressivos (AR) e modelos autorregressivos exógenos (ARX) têm sido amplamente utilizados em aplicações envolvendo sistemas de SHM (Azim & Gül, 2019; Meixedo et al., 2022).

Aliado à utilização de atributos robustos e sensíveis ao dano, é essencial isolar a influência das Variações Ambientais e Operacionais (EOVs), que podem alterar os indicadores. A normalização de atributos torna-se crucial, pois variações ambientais (por exemplo, temperatura) ou operacionais (por exemplo, trens em diferentes velocidades) podem causar maiores variações nos atributos do que aquelas decorrentes de danos. Duas abordagens são comumente citadas na literatura para a normalização: (i) input-output, baseada em modelos de regressão (linear ou não linear), e (ii) output only, baseada em técnicas de variáveis latentes, como a Análise de Componentes Principais (PCA) (Santos et al., 2013). A abordagem input-output mitiga os efeitos das EOVs ao estabelecer relações entre as repostas estruturais medidas e as ações medidas (por exemplo, tráfego, velocidade, temperatura). Em sistemas de monitoramento que não incluem sensores para medir EOVs, técnicas de variáveis latentes podem ser utilizadas, reduzindo a influência das EOVs ao usar apenas medições das repostas estruturais. Normalizações baseadas em PCA, embora tenham se mostrado bastante eficientes, ainda são limitadas ao campo linear.

Mais recentemente, com o significativo avanço das técnicas de aprendizado de máquina autores tem se aproveitado para explorar o potencial de aplicação destas ferramentas em um contexto de SHM. Ferramentas como autoencoders tem sido usada como elementos poderosos para normalização de atributos, sendo capazes de lidar bem inclusive com não linearidades, sendo muito versáteis na eliminação de EOVs dos atributos (Silva et al., 2019) Não só com aplicação à normalização de atributos, modelos de aprendizado profundo têm demonstrado potencial na extração automática de atributos sensíveis ao dano de registros dinâmicos (Ruiz et al., 2025). Recentemente com o advento das técnicas de aprendizado de máquina aplicações de séries temporais tem despertado atenção empregando modelos mais sofisticados, como Redes Neurais Recorrentes (RNN), Unidade Recorrente Gated (GRU) e Memória de Longo Curto Prazo (LSTM), também têm se mostrado promissoras, entretanto sua aplicação ao modal ferroviário ainda é muito incipiente (Ribeiro et al., 2024)

Dentro deste contexto, este trabalho busca investigar a aplicação de modelos transformers em SHM de pontes ferroviárias. Dados de séries temporais coletados por acelerômetros instalados diretamente na estrutura da ponte inteira são utilizados como base de treinamento de um modelo transformer para predição das series temporais destes sensores. Os resíduos destas predições são então utilizados como atributos sensíveis ao dano. A eficácia do modelo foi avaliada por meio de um estudo de caso envolvendo a ponte ferroviária KW51, utilizando dados experimentais para validar a abordagem proposta. Com base neste estudo de caso pode-se demonstrar grande robustez da metodologia na detecção de anomalias.

## 2 Metodologia

### 2.1 Encoder-Only Transformer

A arquitetura clássica de transformer (Vaswani et al., 2017a) com ligeiras alterações será utilizada para detecção de anomalia baseada em resíduo. No entanto, salienta-se que, a complexidade computacional cresce quadraticamente com o aumento da sequência, de tal forma que para aplicações otimizadas e passíveis de aplicação em cenários reais, esparsidade deve ser introduzida (Ilbert et al., 2024; Liu et al., 2024; Wu et al., 2021). A arquitetura do transformer utilizada está representada a seguir.

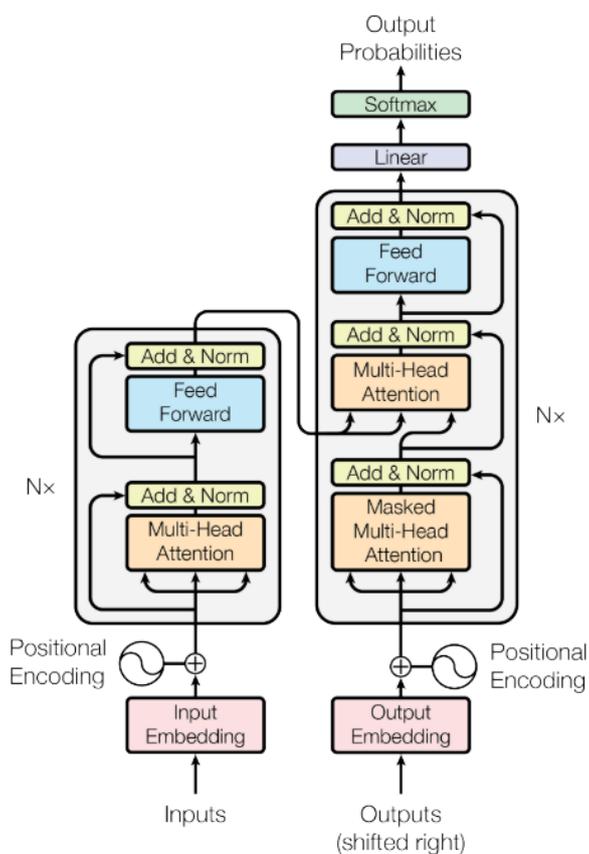


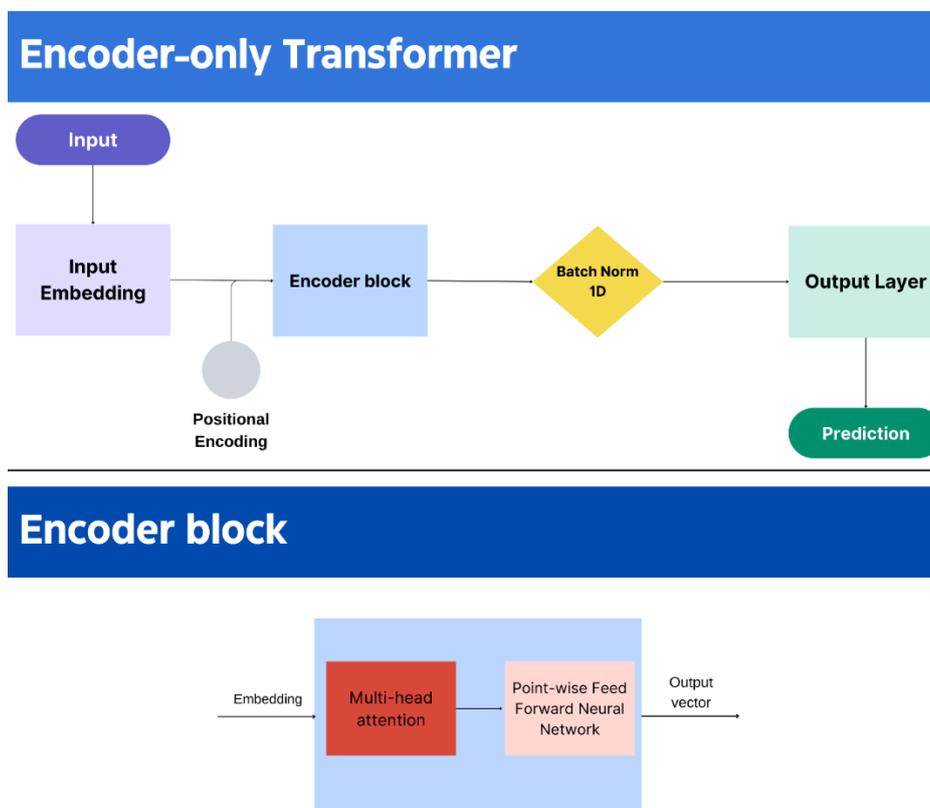
Figura 1- Arquitetura transformer (Vaswani et al., 2017)

A Figura 1 exibe dois blocos, encoder e decoder, que podem ser empilhados  $N$  vezes, conforme representação. A arquitetura pode ser tão profunda quanto necessário, contudo, é preciso observar que o aumento de número de parâmetros acarretará na necessidade de mais dados ou estratégias de regularização mais agressivas (Krizhevsky et al., 2017) para evitar sobreajuste. A ocorrência desse inviabiliza a aplicação em outros cenários senão aqueles do treino, pelo fato do modelo ter aprendido a reproduzir até mesmo os ruídos e interferências, que não serão os mesmos ao variar o cenário.

O bloco encoder é responsável por aprender informações hierárquicas de alto nível utilizando contexto e contribuições individuais dos dados para predição, isto é, através do ajuste dos parâmetros do modelo como pesos e vieses para as diversas camadas, aprender como cada uma das amostras da série temporal em análise se relaciona com os demais pontos (Wu et al., 2021).

O bloco decoder utiliza essas informações para fazer a inferência ou predição, a depender da aplicação desejada. Na utilização original, o bloco decoder viabiliza que a estrutura semântica aprendida seja “transferida” para outros conjuntos, como na interação linguagem-texto, ou entre idiomas (Vaswani et al., 2017). Nesse contexto, a arquitetura utilizada nesse estudo visa apenas prever o próximo valor a partir das representações hierárquicas aprendidas. Dessa forma, será aplicada uma arquitetura encoder-only, representada na Figura a seguir.

A primeira etapa consiste no Input Embedding, em que matrizes de pesos são inicializadas permitindo que em uma dada sequência de entrada diferentes atributos sejam aprendidos automaticamente. Como essa arquitetura é completamente paralelizável, o positional encoding é utilizado para armazenar a ordem dos valores. Essas informações são utilizadas como entrada no bloco de encoder. Nesse bloco, conexões residuais são empregadas por setas semi-circulares, antes da layer Multi-Head Attention até Add and Norm layer, que representa a soma do input antes da camada de atenção com normalização. Essas conexões são usuais em arquitetura profundas, como forma regularização, auxiliando a convergência (He et al., 2015) e garantindo alguma interpretabilidade física (Cross et al., 2022).



**Figura 2 - Arquitetura transformer encoder-only+batch norm layer**

Em sequência, a camada Multi-Head Attention é utilizada para aprender informações contextuais através do mapeamento de um vetor Query e o par Key-Value à um output. Os valores do output dessa camada são calculados pela Equação 1.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Em que  $Q, K$  e  $V$  são os vetores query, key e value,  $d_k$  é a dimensão dos vetores. Essa camada aprende quais valores dos embeddings estão mais fortemente relacionados ao longo da sequência inteira. Esse processamento é completamente paralelizável, porém de alta complexidade (Wu et al., 2021).

O output da camada de atenção é utilizado para uma point-wise Feed-Forward Neural Network que permite que a contribuição particular de cada produto  $Q$  por  $KV$ , seja aprendido. O vetor resultante passa então pela camada de batch normalization, que visa selecionar os features mais proeminentes. Isso é efetuado através da equação:

$$y = \gamma \times \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (2)$$

Em que  $x$  é o valor de entrada (no nosso caso, cada componente do vetor resultante do bloco encoder,  $\mu$  é a média dos valores de  $x$  calculada sobre o batch,  $\sigma$  é a variância calculada sobre o batch,  $\epsilon$  é um valor pequeno para evitar a divisão por zero,  $\gamma$  e  $\beta$  são parâmetros ajustados que permitem escalar os dados no batch. Finalmente, os outputs normalizados são utilizados para fazer a predição.

## 2.2 Discriminação Estatística

Após a predição, a distribuição do erro quadrático (SE) pode ser calculada através da expressão:

$$SE = (y - \hat{y})^2 \quad (3)$$

Em que  $y$  é o valor real e  $\hat{y}$  é o valor da predição. Estudos apontam que, para um modelo cujo treino convergiu, SE tende à uma normal (Wang et al., 2022). Adicionalmente, espera-se que a presença de anomalias interfira nos momentos estatísticos da distribuição (Ribeiro et al., 2024). Dessa forma, o teste t é utilizado para inferir, usando intervalo de confiança de 95%, se essas distribuições são estatisticamente diferentes. Dessa forma as hipóteses são:

- $H_0$ : A diferença entre as duas distribuições é nula;
- $H_1$ : A diferença entre as duas distribuições é não nula

Ao comparar duas distribuições quaisquer, o teste deve inferir a base de dados alterada com diferença não nula para o nível de significância adotado, indicando presença de anomalia. Para este estudo, a comparação é feita antes e pós-intervenção.

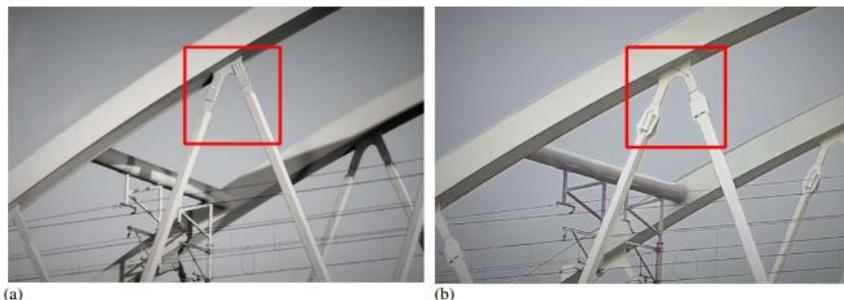
## 2.3 Fluxo de trabalho

A sequência de etapas encontra-se condensada na Figura 3.



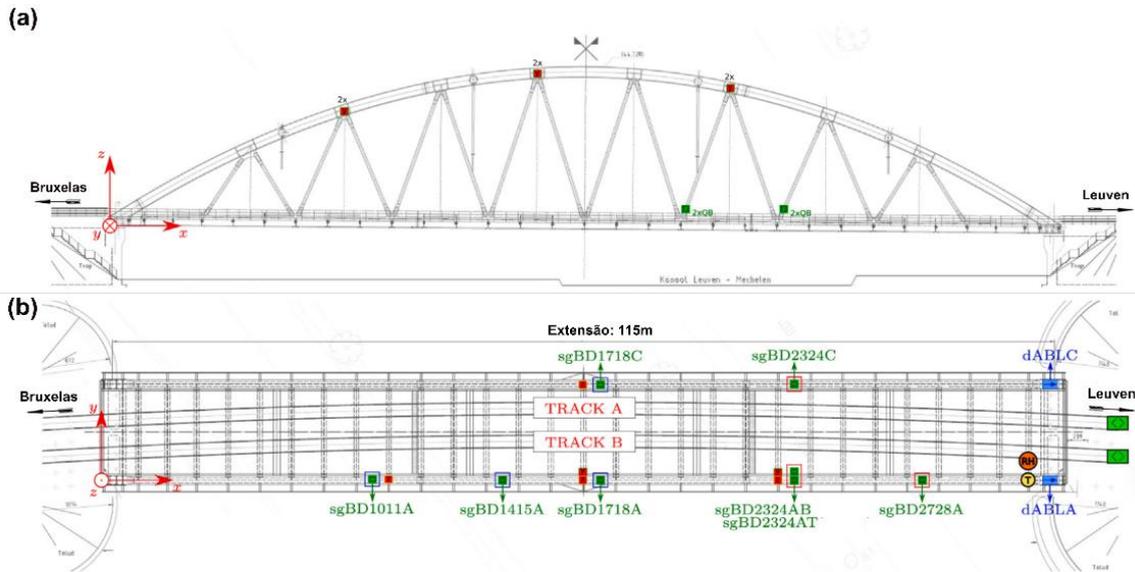
**Figura 3- Fluxo de trabalho**

Inicialmente, serão utilizados como input os dados de aceleração da ponte ferroviária KW51. Após constatada necessidade de trocar as ligações, uma campanha experimental foi iniciada, medindo dados antes, durante e após a alteração global da estrutura. As ligações originais e finais encontram-se representadas na **Figura 4**.



**Figura 4- Ligações: (a) Antes da intervenção (b) Após a alteração**

Os dados utilizados foram de aceleração, já devidamente tratados, conforme referência indicada, coletados por passagem de trem, instalados no tabuleiro. A disposição dos sensores e a ponte em análise estão representados na **Figura 5**, em perfil (4a) e em planta (4b). Os extensômetros estão em verde, os acelerômetros em vermelho, o deslocamento uniaxial em azul, termopares em amarelo em sensores de umidade em laranja. Uma vez que o trabalho intenta focar na aplicação da arquitetura de deep learning, recomenda-se consultar Maes & Lombaert (2021) para maiores detalhes



**Figura 5- Disposição de sensores – (a) Vista em perfil (b) Vista em planta**

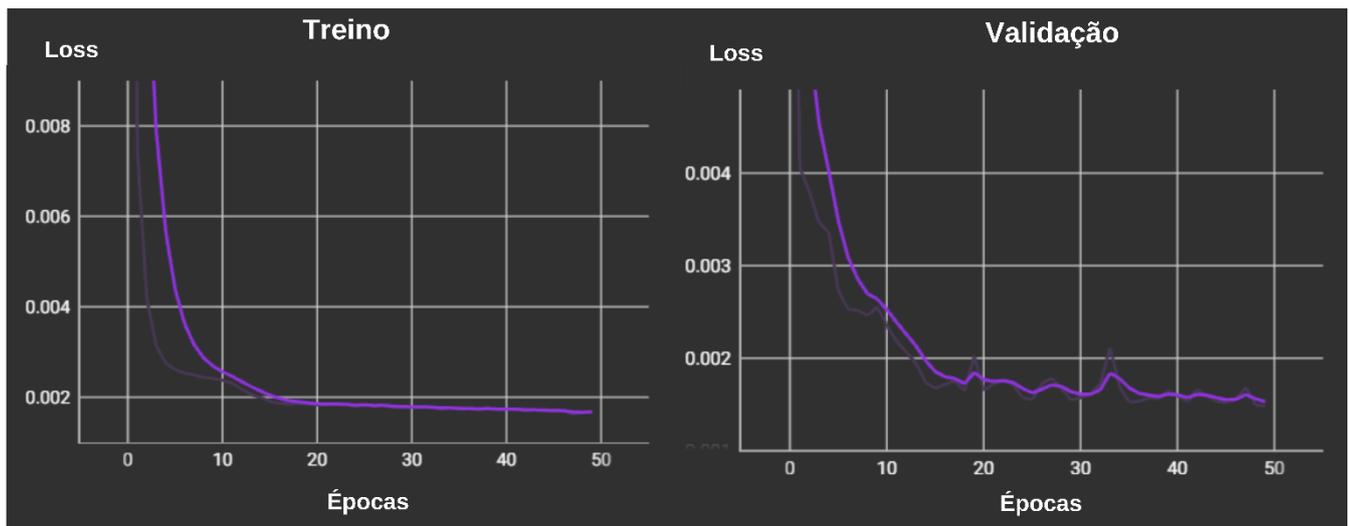
Uma vez que os dados foram coletados ao longo de passagens de um ou mais trens, os dados precisam ser escalados para assegurar a convergência dos algoritmos. Para isso optou-se pela normalização expressa na Equação 4:

$$y_{norm} = \frac{y - y_{min}}{y_{max} - y_{min}} \quad (4)$$

Os dados normalizados são treinados utilizando o modelo descrito na seção 2.1, utilizando um bloco de encoder, sequência de 100 dados, 10 vetores de embeddings e feed-forward de 64x100. Esse processo é efetuado para os dados antes da intervenção e após intervenção para quatro sensores verticais usando 5 passagens para o estado base e para o alterado. Em sequência, procede-se o cálculo de SE e discriminação estatística conforme seção 2.2.

### 3 Resultados

Inicialmente, deve-se atestar a convergência da rede para assegurar os pressupostos de normalidade dos resíduos do teste. A Figura 6 apresenta uma das formas de averiguar isso.

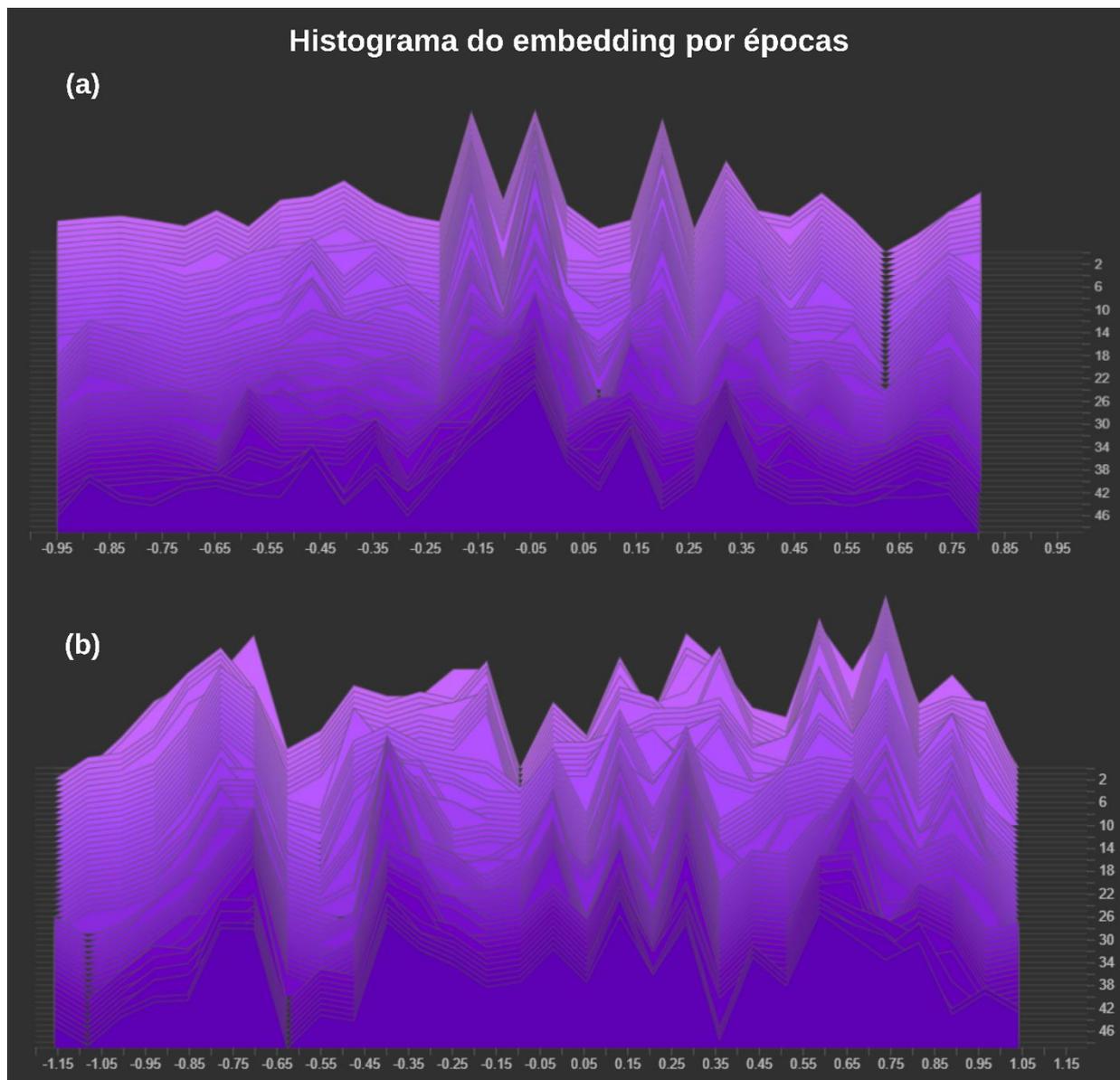


**Figura 6- Curvas de validação para conjunto de treino e de validação**

Percebe-se comportamento de convergência para os conjuntos em torno da mesma região. Contudo, ainda se nota uma pequena assíntota na seção de treino, que por sua vez relaciona-se com as pequenas perturbações do conjunto de validação. Assim, o conjunto de pesos aqui apresentados necessitariam de ajuste de hiperparâmetros como número de épocas ou critérios de otimização para assegurar generalização. Ademais, a métrica utilizada (MSE), é conhecida por penalizar em demasiados pontos que se afastam do comportamento

esperado, fazendo com que modelos temporais se aproximem da média. Dessa forma, uma abordagem possível é a utilização conjunta de métricas como MSE e Similarity Structural Index (SSIM) para garantir um perfil de erros mais regular e apto à detecção de anomalias (Fan et al., 2025).

Em relação aos resultados associados a topologia, enfatizamos os parâmetros das camadas de embeddings, representados na Figura 7a e b para os pesos e vieses do modelo ajustado. Aqui se encontram representados os histogramas (frequência) ao longo das épocas. Percebe-se que alguns picos permanecem estáveis desde as primeiras iterações. Isto aponta que a rede identificou algum atributo relevante para a previsão correta desde as primeiras iterações. Todavia, nos picos menos proeminentes, ainda há alteração ao longo das épocas, corroborando a necessidade de mais épocas para que o modelo consiga aprender os melhores atributos para prever corretamente as respostas da ponte para mais cenários além daqueles de treino.



**Figura 7- Parâmetros das camadas de embeddings - (a) Pesos (b) Vieses**

Finalmente, a Tabela 1 compila os resultados após efetuar o teste-t para a distribuição do erro SE.

**Tabela 1- Resultado por sensor**

| Acelerômetros |          |                 |
|---------------|----------|-----------------|
| Sensor        | Acurácia | Hipótese aceita |
| 11Az          | 100%     | $H_1$           |
| 17Az          | 100%     | $H_1$           |
| 17Cz          | 100%     | $H_1$           |
| 23Az          | 100%     | $H_1$           |

A Tabela demonstra que, utilizando intervalo de confiança de 95%, todas as passagens do estado alterado foram corretamente atribuídas como distintas daquelas comparadas ao estado base, resultando em 100% de acurácia. Salienta-se, contudo, que ante as informações obtidas pela análise da curva de validação e dos parâmetros da rede, torna-se necessário expandir o número de épocas e efetuar outras comparações para evitar, ou delinear claramente sobreajuste a ruídos.

#### 4 Conclusão

Esse trabalho demonstrou a aplicação de uma arquitetura transformer encoder-only com utilização posterior de resíduos para detecção de anomalias para dados de aceleração de uma ponte ferroviária.

A análise dos resultados do treino demonstrou convergência do modelo, premissa essencial para utilização dos resíduos, contudo, também apontou para eventuais problemas de generalização, usuais em arquiteturas profundas. Embora a discriminação estatística tenha atribuído o rótulo adequadamente, recomenda-se o ajuste de hiperparâmetros e eventual aplicação de técnicas de regularização como dropout (Krizhevsky et al., 2017) para assegurar capacidade de generalização.

Ademais, em problemas altamente não lineares, pode-se utilizar estratégias com curvas de validação híbrida, utilizando parâmetros com maiores penalizações como MSE para épocas iniciais com ajuste fino utilizando métricas mais suaves como SSIM (Fan et al., 2025).

#### Referências

- Alves, V., Meixedo, A., Ribeiro, D., Calçada, R., & Cury, A. (2015). Evaluation of the Performance of Different Damage Indicators in Railway Bridges. *Procedia Engineering*, 114, 746–753. <https://doi.org/10.1016/j.proeng.2015.08.020>
- Azim, M. R., & Gül, M. (2019). Damage detection of steel girder railway bridges utilizing operational vibration response. *Structural Control and Health Monitoring*, 26(11), 1–15. <https://doi.org/10.1002/stc.2447>
- ABNT NBR 9452. Inspeção de pontes, viadutos e passarelas de concreto — Procedimento. Associação Brasileira de Normas Técnicas, 2019.
- Carnevale, M., Collina, A., & Peirlinck, T. (2019). A Feasibility Study of the Drive-By Method for Damage Detection in Railway Bridges. *Applied Sciences*, 9(1), 160. <https://doi.org/10.3390/app9010160>
- Chen, B., Zhao, S., & Li, P. (2014). Application of Hilbert-Huang Transform in Structural Health Monitoring: A State-of-the-Art Review. *Mathematical Problems in Engineering*, 2014(1). <https://doi.org/10.1155/2014/317954>
- Cross, E. J., Gibson, S. J., Jones, M. R., Pitchforth, D. J., Zhang, S., & Rogers, T. J. (2022). *Physics-Informed Machine Learning for Structural Health Monitoring* (pp. 347–367). [https://doi.org/10.1007/978-3-030-81716-9\\_17](https://doi.org/10.1007/978-3-030-81716-9_17)
- Fan, J., Chen, M., Gu, Z., Yang, J., Wu, H., & Wu, J. (2025). SSIM over MSE: A new perspective for video anomaly detection. *Neural Networks*, 107115. <https://doi.org/10.1016/j.neunet.2024.107115>
- Farrar, C. R., & Worden, K. (2012). *Structural Health Monitoring*. Wiley. <https://doi.org/10.1002/9781118443118>
- Finotti, R. P., Cury, A. A., & Barbosa, F. de S. (2019). An SHM approach using machine learning and

- statistical indicators extracted from raw dynamic measurements. *Latin American Journal of Solids and Structures*, 16(2), 1–17. <https://doi.org/10.1590/1679-78254942>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. <https://doi.org/https://doi.org/10.48550/arXiv.1512.03385>
- Ilbert, R., Odonnat, A., Feofanov, V., Virmaux, A., Paolo, G., Palpanas, T., & Redko, I. (2024). SAMformer: Unlocking the Potential of Transformers in Time Series Forecasting with Sharpness-Aware Minimization and Channel-Wise Attention. *Proceedings of Machine Learning Research*, 235, 20924–20954.
- Kankanamge, Y., Hu, Y., & Shao, X. (2020). Application of wavelet transform in structural health monitoring. *Earthquake Engineering and Engineering Vibration*, 19(2), 515–532. <https://doi.org/10.1007/s11803-020-0576-8>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). Itransformer: Inverted Transformers Are Effective for Time Series Forecasting. *12th International Conference on Learning Representations, ICLR 2024*, 1–25.
- Maes, K., & Lombaert, G. (2021). *Monitoring Railway Bridge KW51 Before , During , and After Retrofitting*. 26(September 2018), 1–19. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001668](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001668)
- Meixedo, A., Ribeiro, D., Santos, J., Calçada, R., & Todd, M. (2022). Structural health monitoring strategy for damage detection in railway bridges using traffic induced dynamic responses. In *Rail Infrastructure Resilience* (pp. 389–408). Elsevier. <https://doi.org/10.1016/b978-0-12-821042-0.00011-3>
- Meixedo, A., Santos, J., Ribeiro, D., Calçada, R., & Todd, M. (2021). Damage detection in railway bridges using traffic-induced dynamic responses. *Engineering Structures*, 238. <https://doi.org/10.1016/j.engstruct.2021.112189>
- Ribeiro, I., Meixedo, A., Ribeiro, D., & Bittencourt, T. N. (2024). Linear and nonlinear time-series methodologies for bridge condition assessment: A literature review. *Advances in Structural Engineering*, 0(0), 1–24. <https://doi.org/10.1177/13694332241260133>
- Ruiz, D. V., Bragança, C. S. C. de, Poncetti, B. L., Bittencourt, T. N., & Futai, M. M. (2024). Vibration-based structural damage detection strategy using FRFs and machine learning classifiers. *Structures*, 59, 105753. <https://doi.org/10.1016/j.istruc.2023.105753>
- Ruiz, D. V., de Bragança, C. S. C., Poncetti, B. L., Bittencourt, T. N., & Futai, M. M. (2025). Structural damage detection for a small population of nominally equal beams using PSO-optimized Convolutional Neural Networks. *Mechanical Systems and Signal Processing*, 225, 112276. <https://doi.org/10.1016/j.ymsp.2024.112276>
- Santos, J. P., Crémona, C., Orcesi, A. D., & Silveira, P. (2013). Multivariate statistical analysis for early damage detection. *Engineering Structures*, 56, 273–285. <https://doi.org/10.1016/j.engstruct.2013.05.022>
- Silva, M., Santos, A., Santos, R., Figueiredo, E., Sales, C., & Costa, J. C. W. A. (2019). Deep principal component analysis: An enhanced approach for structural damage identification. *Structural Health Monitoring*, 18(5–6), 1444–1463. <https://doi.org/10.1177/1475921718799070>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017a). *Attention Is All You Need*. <http://arxiv.org/abs/1706.03762>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017b). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 5999–6009.
- Wang, C., Ansari, F., Wu, B., Li, S., & Morgese, M. (2022). LSTM approach for condition assessment of suspension bridges based on time-series de flection and temperature data. *Advances in Structural Engineering*, 25(66), 3450–3463. <https://doi.org/10.1177/13694332221133604>
- Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems*, 27(NeurIPS), 22419–22430.